

# New Bounds For Frameproof Codes

Chong Shangguan, Xin Wang, Gennian Ge and Ying Miao

## Abstract

Frameproof codes are used to fingerprint digital data. It can prevent copyrighted materials from unauthorized use. In this paper, we study upper and lower bounds for  $w$ -frameproof codes of length  $N$  over an alphabet of size  $q$ . The upper bound is based on a combinatorial approach and the lower bound is based on a probabilistic construction. Both bounds can improve previous results when  $q$  is small compared to  $w$ , say  $cq \leq w$  for some constant  $c \leq q$ . Furthermore, we pay special attention to binary frameproof codes. We show a binary  $w$ -frameproof code of length  $N$  can not have more than  $N$  codewords if  $N < \binom{w+1}{2}$ .

## Index Terms

Fingerprinting, frameproof code, combinatorial counting, deletion method

## I. INTRODUCTION

Frameproof codes were first introduced by Boneh and Shaw [7] in the context of digital fingerprinting to prevent copyrighted materials from unauthorized use. A set of unique labels, known as digital fingerprints, are inserted into digital data before the distributor wants to sell the copies to different customers. The fingerprints can be viewed as decoder boxes or access control for some copyrighted materials. They are generally denoted as codewords in  $\mathcal{A}^N$ , where  $\mathcal{A}$  is an alphabet of size  $q$  and  $N$  is a positive integer. As long as a coalition of some users want to produce a pirate decoder, they can share and compare their copies. A set of fingerprints is called to be a  $w$ -frameproof code if any coalition of at most  $w$  users can not frame another user not in this coalition.

### A. Previous work

Consider a code  $\mathcal{C} \subseteq \mathcal{A}^N$ . Without loss of generality, we set  $\mathcal{A} = \{1, \dots, q\}$  for  $q \geq 3$  through out this paper. When considering the binary case, we usually set  $\mathcal{A} = \{0, 1\}$ . We call this code an  $(N, n, q)$  code if  $|\mathcal{C}| = n$ . Each codeword  $\vec{c} \in \mathcal{C}$  can be represented as  $\vec{c} = (c_1, \dots, c_N)$ , where  $1 \leq c_i \leq q$  for all  $1 \leq i \leq N$ . For any subset of codewords  $D \subseteq \mathcal{C}$ , we denote  $desc_i(D) = \{c_i : \vec{c} \in D\}$  for every  $1 \leq i \leq N$ . The set of *descendants* of  $D$  is defined as

$$desc(D) = \{\vec{c} \in \mathcal{A}^N : c_i \in desc_i(D), 1 \leq i \leq N\}.$$

Let  $\mathcal{C}$  be an  $(N, n, q)$  code and let  $w \geq 2$  be an integer.  $\mathcal{C}$  is called a  $w$ -frameproof code if we have  $desc(D) \cap \mathcal{C} = D$  for all  $D \subseteq \mathcal{C}$  and  $|D| \leq w$ . In the literature, there are a lot of papers about the properties and applications of frameproof codes, see for example [3], [7], [9], [10], [11], [12], [13], [14], [16]. It is worth mentioning that frameproof codes were widely considered having no traceability for generic digital fingerprinting. However, in [10] the authors showed that frameproof codes have very good traceability for multimedia fingerprinting. There are also many objects related to frameproof codes, such as identifiable parent property codes [1], [4], traceability codes [5] and separating hash families [2], [6], [15].

The determination of upper and lower bounds of frameproof codes is crucial problems in this research area. For upper bounds, when  $q \geq w$  many strong bounds can be found in [3], [12] and [16]. Let  $M_{w,N}(q)$  be the largest cardinality of an  $(N, n, q)$   $w$ -frameproof code and let  $R_{w,N} = \lim_{q \rightarrow \infty} M_{w,N}(q)/q^{\lceil N/w \rceil}$ . It has been determined by Blackburn [3] that  $\lim_{q \rightarrow \infty} \log_q M_{w,N}(q) = \lceil N/w \rceil$ ,  $R_{w,N} = 1$  when  $N \equiv 1 \pmod{w}$ ,  $R_{w,N} = 2$  when  $w = 2$  and  $N$  is even. Existing constructions are usually based on some finite fields with cardinality larger than  $N$ , in other words, the alphabet size  $q$  is larger than the code length  $N$ . Notice that this setting does not work if one wants to know about the code rate defined as  $\alpha_{w,q} = \lim_{N \rightarrow \infty} \frac{1}{N} \log_q M_{w,N}(q)$ , where  $q$  is some given positive integer. When  $q < w$ , much less is known about the upper bounds. In this paper we will concentrate on this theme. For lower bounds, in [15], Stinson, Wei and Chen presented a general result by probabilistic method, where frameproof codes were viewed as a special type of separating hash families.

These bounds are restated as follows.

The research of G. Ge was supported by the National Natural Science Foundation of China under Grant No. 61171198 and Grant No. 11431003, and Zhejiang Provincial Natural Science Foundation of China under Grant No. LZ13A010001. The research of Y. Miao was supported by JSPS Grant-in-Aid for Scientific Research (C) under Grant No. 24540111.

C. Shangguan is with the Department of Mathematics, Zhejiang University, Hangzhou 310027, China (e-mail: 11235061@zju.edu.cn).

X. Wang is with the Department of Mathematics, Zhejiang University, Hangzhou 310027, China (e-mail: 11235062@zju.edu.cn).

G. Ge is with the School of Mathematical Sciences, Capital Normal University, Beijing 100048, China (e-mail: gnge@zju.edu.cn). He is also with Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing, 100048, China.

Y. Miao is with the Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan (miao@sk.tsukuba.ac.jp).

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

**Theorem I.1.** ([3]) Let  $t \in \{1, \dots, w\}$  be an integer such that  $t \equiv N \pmod{w}$ . If there exists an  $(N, n, q)$   $w$ -frameproof code, then

$$n \leq \left( \frac{N}{N - (t-1)\lceil N/w \rceil} \right) q^{\lceil N/w \rceil} + \mathcal{O}(q^{\lceil N/w \rceil - 1}).$$

**Theorem I.2.** ([15]) There exists an  $(N, n, q)$   $w$ -frameproof code provided that

$$n \leq \left(1 - \frac{1}{w!}\right) \left( \frac{q^w}{q^w - (q-1)^w} \right)^{\frac{N}{w}}.$$

**Remark I.1.** In the original paper [15], this bound is stated in the form of separating hash family. Note that a separating hash family of type  $(N, n, q, \{1, w\})$  is equivalent to an  $(N, n, q)$   $w$ -frameproof code.

In a recent paper of Guo, Stinson and Trung [11], the authors paid particular attention to binary frameproof codes with small code length. They showed that if there exists an  $(N, n, 2)$   $w$ -frameproof code  $\mathcal{C}$  with  $w \geq 3$  and  $w+1 \leq N \leq 3w$ , then it always holds that  $n \leq N$ . The equality holds if and only if the representation matrix of  $\mathcal{C}$  in standard form is a permutation matrix of degree  $N$ .

**Theorem I.3.** ([11]) For all  $w \geq 3$  and for all  $w+1 \leq N \leq 3w$ , an  $(N, n, 2)$   $w$ -frameproof code exists only if  $n \leq N$ . If  $n = N$ , then the representation matrix in standard form must be equivalent to a permutation matrix of degree  $N$ .

## B. Main results

In this paper, we mainly investigate the bounds for frameproof codes. We focus on the situation when  $q$  is relatively small compared to  $w$ . There are few results on this case in previous papers. We study the behavior of the code rate  $\alpha_{w,q} = \lim_{N \rightarrow \infty} \frac{1}{N} \log_q M_{w,N}(q)$  as  $N$  approximates infinity. Our upper and lower bounds are both better than previous results when  $w \geq cq$ , where  $c$  is some constant. Given an alphabet size  $q$  and if  $w$  satisfies the above condition, the upper bound of  $\alpha_{w,q}$  can be improved from  $\mathcal{O}(1/w)$  in [3] to  $\mathcal{O}(\log w/w^2)$ . For the lower bound, we show there always exists a code with  $\alpha_{w,q} = \Omega(1/w^2)$ . For binary  $w$ -frameproof code of length  $N$ , we prove that if  $N < \binom{w+1}{2}$ , then it always holds that  $n \leq N$ , which is a significant improvement of Theorem I.3.

This paper is organized as follows. Section 2 is about some definitions and lemmas. In Section 3, we prove our upper bound by a combinatorial counting argument. In Section 4, we present a tight bound for binary  $w$ -frameproof code with code length bounded by  $\binom{w+1}{2}$ . For the lower bound, we present our probabilistic construction in Section 5. Section 6 is about some final discussions and problems.

## II. PRELIMINARIES

We will first give some definitions and notations. In Section I we already presented a definition of frameproof code in terms of the descendant set, which is the most common one in the literature. But in this paper we prefer an alternative definition in terms of distance.

**Definition II.1.** For an  $(N, n, q)$  code  $\mathcal{C}$ , the distance between any codeword  $\vec{c} \in \mathcal{C}$  and any collection of codewords  $D \subseteq \mathcal{C}$  is defined as follows:

$$d(\vec{c}, D) = |\{i : 1 \leq i \leq N, c_i \notin \text{desc}_i(D)\}|.$$

$\mathcal{C}$  is called a  $w$ -frameproof code if  $d(\vec{c}, D) > 0$  for all  $|D| \leq w$  and  $\vec{c} \notin D$ .

It is easy to verify that this definition is equivalent to the original one. It is interesting that our ideas in this paper are indeed implied in this distance-based definition. Recall that a codeword  $\vec{c} \in \mathcal{C}$  is denoted as  $\vec{c} = (c_1, \dots, c_N)$ , where  $1 \leq c_i \leq q$  for all  $1 \leq i \leq N$ . We can associate each codeword  $\vec{c}$  with a set of two tuples  $F_{\vec{c}} = \{(i, c_i) : 1 \leq i \leq N\}$ . Note that  $F_{\vec{c}_1} = F_{\vec{c}_2}$  if and only if  $\vec{c}_1 = \vec{c}_2$  and it always holds that  $|F_{\vec{c}}| = N$ . We call  $F_{\vec{c}}$  the set corresponding to  $\vec{c}$ . It is easy to show  $|\cup_{\vec{c} \in \mathcal{C}} F_{\vec{c}}| \leq qN$ . The following lemma establishes an immediate connection between the distance defined in Definition II.1 and the cardinality of  $F_{\vec{c}}$ .

**Lemma II.1.** Suppose  $D \subseteq \mathcal{C}$  is a set of codewords. For every  $\vec{c} \in \mathcal{C}$ , let  $F_{\vec{c}} = \{(i, c_i) : 1 \leq i \leq N\}$  be its corresponding set. Then we have

$$d(\vec{c}, D) = |F_{\vec{c}} \setminus \cup_{\vec{d} \in D} F_{\vec{d}}|.$$

*Proof:* Note that  $d(\vec{c}, D) = |\{i : 1 \leq i \leq N, c_i \notin \text{desc}_i(D)\}|$ . The lemma follows from the simple fact that  $c_i \in \text{desc}_i(D)$  if and only if  $(i, c_i) \in F_{\vec{d}}$  for some  $\vec{d} \in D$ . ■

**Remark II.1.** One can verify that  $\mathcal{C}$  is a  $w$ -frameproof code if and only if  $|F_{\vec{c}} \setminus \cup_{\vec{d} \in D} F_{\vec{d}}| > 0$  for all  $D \subseteq \mathcal{C}$ ,  $|D| \leq w$  and  $\vec{c} \notin D$  directly from Definition II.1.

For a set  $S \subseteq \{1, \dots, N\}$  and  $|S| = t$ , we call  $T = \{(i, c_i) : i \in S\} \subseteq F_{\vec{c}}$  a  $t$ -pattern of  $\vec{c}$  (or equivalently,  $F_{\vec{c}}$ ).  $T$  is said to be an own  $t$ -pattern of  $\vec{c}$  if  $T \not\subseteq F_{\vec{d}}$  for all  $\vec{d} \in \mathcal{C} \setminus \{\vec{c}\}$ . Denote  $\mathcal{C}_t = \{\vec{c} \in \mathcal{C} : \vec{c} \text{ has at least one own } t\text{-pattern}\}$  and  $\mathcal{H}_t = \mathcal{C} - \mathcal{C}_t$ . We have the following two easy lemmas.

**Lemma II.2.**  $|\mathcal{C}_t| \leq \binom{N}{t} q^t$ .

*Proof:* Just notice that every codeword  $\vec{c} \in \mathcal{C}_t$  is uniquely identified by its arbitrary own  $t$ -pattern. ■

**Lemma II.3.** Suppose  $\vec{c} \in \mathcal{H}_t$ . Then for any  $1 \leq j \leq w$  and any distinct  $\vec{c}_1, \vec{c}_2, \dots, \vec{c}_j \in \mathcal{C} \setminus \{\vec{c}\}$ . We have

$$d(\vec{c}, \{\vec{c}_1, \vec{c}_2, \dots, \vec{c}_j\}) = |F_{\vec{c}} \setminus \cup_{1 \leq i \leq j} F_{\vec{c}_i}| \geq (w - j)t + 1.$$

*Proof:* Assume not. Then  $|F_{\vec{c}} \setminus \cup_{1 \leq i \leq j} F_{\vec{c}_i}| \leq (w - j)t$  holds for some  $\vec{c}_1, \dots, \vec{c}_j \in \mathcal{C} \setminus \{\vec{c}\}$ . Note that  $\vec{c} \in \mathcal{H}_t$ , which means any  $t$ -pattern of  $F_{\vec{c}}$  is not an own  $t$ -pattern, that is, it belongs to some  $F_{\vec{d}}$  with  $\vec{d} \in \mathcal{C} \setminus \{\vec{c}\}$ . Then there exist  $w - j$  codewords  $\vec{c}_{j+1}, \dots, \vec{c}_w \in \mathcal{C} \setminus \{\vec{c}\}$  such that  $F_{\vec{c}} \setminus \cup_{1 \leq i \leq j} F_{\vec{c}_i} \subseteq \cup_{j+1 \leq i \leq w} F_{\vec{c}_i}$ . It implies  $|F_{\vec{c}} \setminus \cup_{1 \leq i \leq w} F_{\vec{c}_i}| = 0$ , which contradicts the definition of  $w$ -frameproof code. ■

### III. THE UPPER BOUND

In this section we present our upper bound. For an  $(N, n, q)$   $w$ -frameproof code  $\mathcal{C}$ , we study the behavior of the code rate defined as  $\alpha_{w,q} = \lim_{N \rightarrow \infty} \frac{1}{N} \log_q |\mathcal{C}|$ . Given an alphabet size  $q$ , in [3] the author proved that  $\alpha_{w,q} = \mathcal{O}(1/w)$ . Here we show  $\alpha_{w,q} = \mathcal{O}(\log w/w^2)$ . Our bound is always better if  $w \geq cq$  for some constant  $c$ . The following theorem is based on previous lemmas with a particular observation of  $|\cup_{\vec{c} \in \mathcal{C}} F_{\vec{c}}| \leq qN$ .

**Theorem III.1.** Suppose  $\mathcal{C} \subseteq \mathcal{A}^N$  is an  $(N, n, q)$   $w$ -frameproof code. Then we have

$$|\mathcal{C}| \leq q^{\lceil N(q-1)/\binom{w}{2} \rceil \log_q eq \binom{w}{2}/(q-1)} + w.$$

*Proof:* Suppose  $|\mathcal{H}_t| \geq w + 1$ . Then for distinct  $\vec{c}_1, \dots, \vec{c}_{w+1} \in \mathcal{H}_t$  and the corresponding  $F_{\vec{c}_i}$ , by Lemma II.3 one has  $|\cup_{1 \leq i \leq w+1} F_{\vec{c}_i}| = |F_{\vec{c}_1}| + |F_{\vec{c}_2} \setminus F_{\vec{c}_1}| + \dots + |F_{\vec{c}_{w+1}} \setminus (F_{\vec{c}_1} \cup \dots \cup F_{\vec{c}_w})| \geq N + w + \sum_{j=1}^w (w - j)t = N + w + t \binom{w}{2}$ . The right hand side of above formula exceeds  $qN$  for  $t > \frac{N(q-1)-w}{\binom{w}{2}}$ , which violates the fact  $|\cup_{1 \leq i \leq w+1} F_{\vec{c}_i}| \leq |\cup_{\vec{c} \in \mathcal{C}} F_{\vec{c}}| \leq qN$ . This implies  $|\mathcal{H}_t| \leq w$  for those  $t > \frac{N(q-1)-w}{\binom{w}{2}}$ . Let  $t = \lceil \frac{N(q-1)}{\binom{w}{2}} \rceil$ . Then we get

$$\begin{aligned} |\mathcal{C}| &= |\mathcal{C}_t| + |\mathcal{H}_t| \\ &\leq |\mathcal{C}_t| + w \leq \binom{N}{t} q^t + w \\ &= \left( \frac{N}{\lceil \frac{N(q-1)}{\binom{w}{2}} \rceil} \right) q^{\lceil \frac{N(q-1)}{\binom{w}{2}} \rceil} + w \\ &< \left( \frac{eqN}{\lceil \frac{N(q-1)}{\binom{w}{2}} \rceil} \right)^{\lceil \frac{N(q-1)}{\binom{w}{2}} \rceil} + w \\ &\leq q^{\lceil \frac{N(q-1)}{\binom{w}{2}} \rceil \log_q eq \binom{w}{2}/(q-1)} + w \end{aligned}$$

from Lemma II.2 using  $\binom{N}{t} < (eN/t)^t$ . ■

**Remark III.1.** Compared with Theorem I.1, our bound is an improvement when  $\lceil \frac{N(q-1)}{\binom{w}{2}} \rceil \log_q eq \binom{w}{2}/(q-1) < \lceil N/w \rceil$ . For positive real numbers  $x$  and  $y > 1$ , it can be verified that  $\lceil \frac{x}{y} \rceil < \lceil x \rceil \frac{1}{y-1}$  if  $x$  is sufficiently large, for example,  $x > y^2$ . Thus it can be verified that our code rate is better than Theorem I.1 when  $w \geq 14q$  with  $q \geq 14$ . For  $2 \leq q \leq 14$ , in Table I we list the conditions when our bound is better.

TABLE I: Given  $q$ , the minimal  $w$  that our bound is better

$q$	2	3	4	5
min $w$	25	33	42	51
$q$	6	7	8	9
min $w$	51	60	68	77
$q$	10	11	12	13
min $w$	94	102	110	118

If the alphabet size  $q$  is given, the previous known results often give an upper bound of code rate as  $\alpha_{w,q} = \mathcal{O}(1/w)$ . But our bound is  $\alpha_{w,q} = \mathcal{O}(\log w/w^2) = \mathcal{O}(1/w^{2-\epsilon})$  where  $\epsilon$  is some small quantity related to  $w$ . This difference is quite

reasonable since in the literature there exists good constructions only when  $q > N$ , using some finite fields with size larger than  $N$ . If  $q$  is relatively small, there exists no such good constructions and the explicit upper bound is still far from known. In Section V, we will present a probabilistic construction with  $\alpha_{w,q} = \Omega(1/w^2)$ , which implies  $c_1/w^2 \leq \alpha_{w,q} \leq c_2/w^{2-\epsilon}$  for given  $q$  and sufficiently large  $w$ , where  $c_1$  and  $c_2$  are positive real numbers.

#### IV. TIGHT BOUND ON SMALL CODE LENGTH

In a recent paper of Guo, Stinson and Trung [11], the authors considered the condition when a binary  $w$ -frameproof code of length  $N$  can have more than  $N$  codewords. They proved that if  $\mathcal{C}$  is an  $(N, n, 2)$   $w$ -frameproof code with  $w + 1 \leq N \leq 3w$  and  $w \geq 3$ , then  $n \leq N$ . We will show  $n \leq N$  still holds even when  $N < \binom{w+1}{2}$ .

We begin with some definitions which are originally from [11]. We can depict an  $(N, n, q)$  code as an  $N \times n$  matrix on  $q$  symbols, where each column of the matrix corresponds to one of the codewords. This matrix is called the representation matrix of the code. Consider the representation matrix of any frameproof code. If we permute the entries in each row separately, i.e. a permutation of the elements  $\{1, \dots, q\}$ , we get new frameproof codes which can be considered to be in the same equivalence class with the original one. For binary code, we say it is in *standard form* if every row of its representation matrix has at most  $n/2$  entries of 1.

Recall that  $\mathcal{C} = \mathcal{C}_t \cup \mathcal{H}_t$ . The following theorem is proved by considering the situation of  $t = 1$ , in other words, the 1-patterns.

**Theorem IV.1.** *Suppose  $\mathcal{C}$  is an  $(N, n, 2)$   $w$ -frameproof code with  $w \geq 2$  and  $N < \binom{w+1}{2}$ . Then it always holds that  $n \leq N$ . The equality holds if and only if the representation matrix of  $\mathcal{C}$  in standard form is equivalent to a permutation matrix of order  $N$ .*

In order to present our proof, we would like to give some lemmas first.

**Lemma IV.2.** *Suppose  $\mathcal{C}$  is an  $(N, n, 2)$   $w$ -frameproof code with  $N \leq w$ . Then it always holds that  $\mathcal{C} = \mathcal{C}_1$ .*

*Proof:* The conclusion holds since if there is any  $\vec{c} \in \mathcal{H}_1$ , then for each  $1 \leq i \leq N$  we can find some  $\vec{d}_i \in \mathcal{C} \setminus \{\vec{c}\}$  such that  $c_i = d_i$ , implying  $d(\vec{c}, \{\vec{d}_1, \dots, \vec{d}_N\}) = 0$ . ■

**Lemma IV.3.** *Suppose  $\mathcal{C}$  is an  $(N, n, 2)$   $w$ -frameproof code with  $\mathcal{C} = \mathcal{C}_1$ . Then it always holds that  $n \leq N$ . The equality holds if and only if the representation matrix of  $\mathcal{C}$  in standard form is equivalent to a permutation matrix of order  $N$ .*

*Proof:* If  $\mathcal{C} = \mathcal{C}_1$ , then for every codeword  $\vec{c} \in \mathcal{C}$ , there exists some  $i$ ,  $1 \leq i \leq N$ , such that  $c_i \neq d_i$  for all  $\vec{d} \in \mathcal{C} \setminus \{\vec{c}\}$ . Consider the representation matrix of  $\mathcal{C}$ . Since each  $c_i \in \{0, 1\}$ , then we can set this special  $c_i$  to be 1 and other entries in row  $i$  to be 0. Thus  $n \leq N$  since there is at least one such row for each  $\vec{c} \in \mathcal{C}_1$ . The equality holds if and only if there is exactly one such row for each  $\vec{c} \in \mathcal{C}_1$ . The resulting representation matrix is a permutation matrix of order  $N$ . ■

Now we can prove Theorem IV.1.

*Proof:* The case  $N \leq w$  follows from Lemma IV.2 and Lemma IV.3. If  $N \geq w + 1$ , then we apply induction on  $N$ . The proof is divided into three parts.

- 1) If  $\mathcal{C}_1 = \emptyset$ , then  $\mathcal{C} = \mathcal{H}_1$ . Assume that  $|\mathcal{H}_1| \geq w + 1$ . Then we can choose  $w + 1$  distinct codewords  $\vec{c}_1, \dots, \vec{c}_{w+1} \in \mathcal{H}_1$ . By setting  $t = 1$  in Lemma II.3, we have  $2N \geq |\cup_{1 \leq i \leq w+1} F_{\vec{c}_i}| = |F_{\vec{c}_1}| + |F_{\vec{c}_2} \setminus F_{\vec{c}_1}| + \dots + |F_{\vec{c}_{w+1}} \setminus (F_{\vec{c}_1} \cup \dots \cup F_{\vec{c}_w})| \geq N + w + \sum_{j=1}^w (w - j) = N + w + \binom{w}{2} = N + \binom{w+1}{2}$ , which violates the assumption  $N < \binom{w+1}{2}$ . Therefore,  $|\mathcal{H}_1| \leq w$ . Then the conclusion follows from  $N \geq w + 1$ .
- 2) If  $\mathcal{C}_1 \neq \emptyset$  and  $\vec{c} \in \mathcal{C}_1$ . Then there exists some  $i$ ,  $1 \leq i \leq N$ , such that  $c_i \neq d_i$  for all  $\vec{d} \in \mathcal{C} \setminus \{\vec{c}\}$ . Consider the standard representation matrix of  $\mathcal{C}$ . Since  $c_i \in \{0, 1\}$ , then there is exactly one 1 in the  $i$ -th row of this matrix, which implies  $c_i = 1$ . By permuting the row entries, we can set the other entries of the column indexed by  $\vec{c}$  to be zero. Let  $i = 1$  for simplicity. Then by above discussion, we actually get an equivalent code that has a codeword  $\vec{c} = (1, 0, \dots, 0)$ . The representation matrix is of the following form,

$$\begin{array}{ccccccc} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & * & * & \dots & * & * \\ & & & \dots & & \\ 0 & * & * & \dots & * & * \end{array}$$

where  $*$  denotes a symbol in  $\{0, 1\}$ . By deleting the first row and first column, we can get a new matrix which is the representation matrix of an  $(N - 1, n - 1, 2)$   $w$ -frameproof code. Let us denote this code by  $\mathcal{C}^{(1)}$ .

- 3) Repeat the procedures in step 2 for  $\mathcal{C}^{(1)}$ . If we always have  $\mathcal{C}_1^{(i)} \neq \emptyset$  as  $i$  grows, then we will end up with a matrix where any row contains at most one 1, which implies that  $n \leq N$ , and the equality holds if and only if this matrix is the standard representation matrix of the original code  $\mathcal{C}$ . Otherwise, there is some  $\mathcal{C}^{(i)}$  such that the procedure ends up with some  $\mathcal{C}_1^{(i)} = \emptyset$ , that is,  $\mathcal{C}^{(i)} = \mathcal{H}_1^{(i)}$ . By Lemma IV.2 the code length of  $\mathcal{C}^{(i)}$  is at least  $w + 1$ . Note that  $\mathcal{C}^{(i)}$  is an  $(N - i, n - i, 2)$   $w$ -frameproof code, then the conclusion follows from the arguments in step 1.

Denote  $N(w)$  the minimal  $N$  such that there exists an  $(N, n, 2)$   $w$ -frameproof code with  $n > N$ . It was proved that  $N(w) \geq w + 1$  in [3]. In [11], the authors actually verified  $N(2) = 3$  and  $N(w) \geq 3w + 1$  for  $w \geq 3$ . In this paper it is improved to  $N(w) \geq \binom{w+1}{2}$  for all  $w \geq 2$ . In the following we will present an example that gives rise to  $N(w) \leq w^2 + o(w^2)$ .

**Definition IV.1.** An affine plane is an incidence system of points and lines such that

- (AP1) For any two distinct points, there is exactly one line through both points.
- (AP2) Given any line  $l$  and any point  $P$  not on  $l$ , there is exactly one line through  $P$  that does not meet  $l$ .
- (AP3) There exist four points such that no three are collinear.

For a detailed introduction of affine plane, the readers are referred to [8]. In an affine plane, any two lines have the same number of points, finite or infinite. The order of an affine plane is the number of points on any given line of the plane. If  $\mathcal{P}$  is an affine plane of finite order  $r$ , then it is proved that every point of  $\mathcal{P}$  lies on exactly  $r + 1$  lines and  $\mathcal{P}$  has exactly  $r^2$  points and  $r^2 + r$  lines. Let  $M$  be the incidence matrix of  $\mathcal{P}$ . Then  $M$  is an  $r^2 \times (r^2 + r)$  matrix whose rows are indexed by points of  $\mathcal{P}$  and columns are indexed by lines of  $\mathcal{P}$ . The entry  $M(P, l) = 1$  if point  $P$  is on line  $l$  and  $M(P, l) = 0$  otherwise. By property (AP1), any two columns of  $M$  can agree with at most one 1 in their coordinates. Consider the binary frameproof code with representation matrix  $M$ . It is easy to show this code is an  $(r^2, r^2 + r, 2)$   $(r - 1)$ -frameproof code. For every prime power, it is known that there exists an affine plane of this order. Let  $q$  be the smallest prime power no less than  $w + 1$ . Then the existence of an affine plane of order  $q$  gives rise to a  $(q^2, q^2 + q, 2)$   $w$ -frameproof code. Together with Theorem IV.1 we have the following corollary

**Corollary IV.4.** Let  $N(w)$  be the minimal  $N$  such that there exists an  $(N, n, 2)$   $w$ -frameproof code with  $n > N$ . Then

$$\left(\frac{1}{2} + o(1)\right)w^2 \leq N(w) \leq (1 + o(1))w^2.$$

## V. THE LOWER BOUND

In this section, we use the standard probabilistic method to give an existence result for frameproof codes. The technique we employ is commonly termed the deletion method.

**Theorem V.1.** If  $q \leq w + 1$ , then there exists an  $(N, n, q)$   $w$ -frameproof code with cardinality at least

$$\frac{1}{2^{\frac{w+1}{w}}} \left( \frac{1}{1 - \left(1 - \frac{q-1}{w+1}\right)^w - \frac{q-1}{w+1} \left(\frac{w}{w+1}\right)^w} \right)^{\frac{N}{w}}.$$

*Proof:* Assume  $\mathcal{A} = \{1, \dots, q\}$  is an alphabet of size  $q$ . Then choose  $2M$  vectors in  $\mathcal{A}^N$  independently with repetitions. Each vector  $\vec{c} = (c_1, \dots, c_N)$  is generated as follows. For every  $1 \leq i \leq N$ , we set  $\Pr[c_i = q] = \lambda$  and  $\Pr[c_i = j] = \mu$  for each  $1 \leq j \leq q - 1$ . Obviously it holds that  $\lambda + (q - 1)\mu = 1$ . The values  $M$ ,  $\lambda$  and  $\mu$  will be determined later. Denote the obtained random family by  $\mathcal{C}_0$ . For some  $\vec{c} \in \mathcal{C}_0$  and arbitrary  $D \subseteq \mathcal{C}_0 \setminus \{\vec{c}\}$  with  $|D| = w$ , we compute the probability that  $\vec{c}$  and  $D$  violate the condition of  $w$ -frameproof code, i.e.,  $\Pr[d(\vec{c}, D) = 0]$ . We have

$$\Pr[d(\vec{c}, D) = 0] = \prod_{i=1}^N \Pr[c_i \in \text{desc}_i(D)],$$

where

$$\Pr[c_i \in \text{desc}_i(D)] = \lambda(1 - (1 - \lambda)^w) + (q - 1)\mu(1 - (1 - \mu)^w)$$

holds for every  $1 \leq i \leq N$ . Denote  $\Pr[c_i \in \text{desc}_i(D)] := P(\lambda, \mu)$  for convenience. Then  $\Pr[d(\vec{c}, D) = 0] = (P(\lambda, \mu))^N$ . Thus for  $M < 2^{-\frac{w+1}{w}} P(\lambda, \mu)^{-\frac{N}{w}}$ , the expected number of pairs  $\vec{c}$  and  $D \subseteq \mathcal{C}_0 \setminus \{\vec{c}\}$  violating the property of  $w$ -frameproof code is bounded by

$$\begin{aligned} &\leq \binom{2M}{1} \binom{2M-1}{w} \Pr[d(\vec{c}, D) = 0] \\ &< (2M)^{w+1} (P(\lambda, \mu))^N < M, \end{aligned}$$

that is, there exists a code  $\mathcal{C}_0$  of cardinality  $|\mathcal{C}_0| = 2M$  with at most  $M$  pairs  $\vec{c}$  and  $D \subseteq \mathcal{C}_0 \setminus \{\vec{c}\}$ ,  $|D| = w$ , violating the property of  $w$ -frameproof code.

Fix such a code  $\mathcal{C}_0$  and for each of the pairs  $(\vec{c}, D)$  delete the inadmissible vector  $\vec{c}$ . Denote the resulting code by  $\mathcal{C}$ . Then the expected number of the remaining vectors in  $\mathcal{C}$  is greater than  $2M - M = M$ , and the vectors in  $\mathcal{C}$  satisfy the condition of  $w$ -frameproof code. It concludes that we have shown the existence of a  $w$ -frameproof code  $\mathcal{C}$  with at least  $M$  codewords. Then the theorem follows by setting  $\lambda = 1 - \frac{q-1}{w+1}$  and  $\mu = \frac{1}{w+1}$ .  $\blacksquare$

**Remark V.1.** Assume  $p_j$  is the probability that  $\Pr[c_i = j] = p_j$  for  $1 \leq i \leq N$  and each  $j \in \{1, \dots, q\}$ . The core of our construction is in fact the optimization

$$\begin{aligned} \max \quad & \sum_{j=1}^q p_j (1 - p_j)^w \\ \text{s.t.} \quad & \sum_{j=1}^q p_j = 1. \end{aligned}$$

In [15], they just choose  $p_1 = p_2 = \dots = p_q = \frac{1}{q}$ . Our choice is better when  $q \leq w + 1$ .

**Remark V.2.** Compared with the bound in Theorem I.2, it is easy to see our bound on code rate is an improvement if  $(1 - \frac{q-1}{w+1})(\frac{q-1}{w+1})^w + \frac{q-1}{w+1}(\frac{w}{w+1})^w > (\frac{q-1}{q})^w$ . One can verify that the above inequality holds when  $q \leq \frac{w}{2} + 1$  if  $w \geq 8$  (see Appendix). For small  $w$  and  $q$ , in Table 2 we list the conditions when the inequality is satisfied.

TABLE II: Given  $q$ , the minimal  $w$  that our bound is better

$q$	2	3	4
min $w$	5	7	8
$q$	5	40	41
min $w$	8	49	50

Last two values indicate that the ratio of  $q/w$  may approximate 1 as  $q$  grows. We choose  $q/w = 1/2$  just for the sake of easy computing. By Taylor expansion of the function  $\ln(1+x)$  one can show the rate of our lower bound is  $\Omega(1/w^2)$ . There is still a gap against the upper bound  $\mathcal{O}(\log w/w^2)$  determined in Theorem III.1. To see how better our bound is than [15] when  $q$  is fixed and  $w$  grows large, we just present both bounds for  $q = 2$ , which are  $\Omega(1/w2^w)$  and  $\Omega(1/w^2)$  respectively.

## VI. CONCLUSIONS

In this paper we investigate the bounds of frameproof codes. To prove the upper bound we use a pure combinatorial approach. For the lower bound, we use a probabilistic method. Our main ideas are from the distance implied in the structure of frameproof codes. Compared with previous bounds, it is easy to find out that the cardinality of frameproof codes differs widely according to the numerical relationship of frameproof property  $w$ , alphabet size  $q$  and code length  $N$ . For large  $q$ , the authors of [3] gave several good constructions that fit in the upper bound. But it is not the case for small  $q$ , where the gap between the lower and the upper bound is still huge. It is nice if one can construct frameproof codes on small alphabet with good code rate. For binary  $w$ -frameproof code of length  $N$ , people pay particular attention to the extremal situation that one can find a code having at least  $N + 1$  codewords. In this paper we show the necessary condition is  $N \geq \binom{w+1}{2}$ . And it is showed that the sufficient condition is  $N = w^2 + o(w^2)$ . Any improvement of this result may be interesting.

## APPENDIX

We want to show  $(1 - \frac{q-1}{w+1})(\frac{q-1}{w+1})^w + \frac{q-1}{w+1}(\frac{w}{w+1})^w > (\frac{q-1}{q})^w$  when  $q \leq w/2 + 1$  and  $w \geq 8$ . It suffices to show  $\frac{q-1}{w+1}(\frac{w}{w+1})^w > (\frac{q-1}{q})^w$ , which is equivalent to  $\frac{w^w}{(w+1)^{w+1}} > \frac{(q-1)^{w-1}}{q^w}$ . Note that  $f(x) = \frac{(x-1)^{w-1}}{x^w}$  is monotonically increasing when  $1 < x < w$ . Therefore,  $\frac{(q-1)^{w-1}}{q^w} < \frac{(w/2)^{w-1}}{(w/2+1)^w}$ . It suffices to show  $\frac{w^w}{(w+1)^{w+1}} > \frac{2w^{w-1}}{(w+2)^w}$ , which is equivalent to  $(1 + \frac{1}{w+1})^{w+1} > \frac{2(w+2)}{w}$ . Note that the left side of the last inequality is monotonically increasing and right side is monotonically decreasing. By direct computation one can show the inequality holds if  $w \geq 8$ .

## REFERENCES

- [1] N. Alon and U. Stav, "New bounds on parent-identifying codes: the case of multiple parents," *Combin. Probab. Comput.*, vol. 13, no. 6, pp. 795–807, 2004. [Online]. Available: <http://dx.doi.org/10.1017/S0963548304006388>
- [2] M. Bazrafshan and T. Trung, "Bounds for separating hash families," *J. Combin. Theory Ser. A*, vol. 118, no. 3, pp. 1129–1135, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.jcta.2010.11.006>
- [3] S. R. Blackburn, "Frameproof codes," *SIAM J. Discrete Math.*, vol. 16, no. 3, pp. 499–510 (electronic), 2003. [Online]. Available: <http://dx.doi.org/10.1137/S0895480101384633>
- [4] —, "An upper bound on the size of a code with the  $k$ -identifiable parent property," *J. Combin. Theory Ser. A*, vol. 102, no. 1, pp. 179–185, 2003. [Online]. Available: [http://dx.doi.org/10.1016/S0097-3165\(03\)00030-X](http://dx.doi.org/10.1016/S0097-3165(03)00030-X)
- [5] S. R. Blackburn, T. Etzion, and S. Ng, "Traceability codes," *J. Combin. Theory Ser. A*, vol. 117, no. 8, pp. 1049–1057, 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.jcta.2010.02.009>
- [6] S. R. Blackburn, T. Etzion, D. R. Stinson, and G. M. Zaverucha, "A bound on the size of separating hash families," *J. Combin. Theory Ser. A*, vol. 115, no. 7, pp. 1246–1256, 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.jcta.2008.01.009>
- [7] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1897–1905, 1998. [Online]. Available: <http://dx.doi.org/10.1109/18.705568>
- [8] F. Buekenhout, Ed., *Handbook of incidence geometry*. North-Holland, Amsterdam, 1995, buildings and foundations.
- [9] Y. M. Chee and X. Zhang, "Improved constructions of frameproof codes," *IEEE Trans. Inform. Theory*, vol. 58, no. 8, pp. 5449–5453, 2012. [Online]. Available: <http://dx.doi.org/10.1109/TIT.2012.2197812>

- [10] M. Cheng and Y. Miao, "On anti-collusion codes and detection algorithms for multimedia fingerprinting," *Information Theory, IEEE Transactions on*, vol. 57, no. 7, pp. 4843–4851, July 2011.
- [11] C. Guo, D. R. Stinson, and T. Trung, "On tight bounds for binary frameproof codes," <http://arxiv.org/abs/1406.6920>, 2014.
- [12] J. N. Staddon, D. R. Stinson, and R. Wei, "Combinatorial properties of frameproof and traceability codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1042–1049, 2001. [Online]. Available: <http://dx.doi.org/10.1109/18.915661>
- [13] D. R. Stinson, T. Trung, and R. Wei, "Secure frameproof codes, key distribution patterns, group testing algorithms and related structures," *J. Statist. Plann. Inference*, vol. 86, no. 2, pp. 595–617, 2000, special issue in honor of Professor Ralph Stanton. [Online]. Available: [http://dx.doi.org/10.1016/S0378-3758\(99\)00131-7](http://dx.doi.org/10.1016/S0378-3758(99)00131-7)
- [14] D. R. Stinson and R. Wei, "Combinatorial properties and constructions of traceability schemes and frameproof codes," *SIAM J. Discrete Math.*, vol. 11, no. 1, pp. 41–53 (electronic), 1998. [Online]. Available: <http://dx.doi.org/10.1137/S0895480196304246>
- [15] D. R. Stinson, R. Wei, and K. Chen, "On generalized separating hash families," *J. Combin. Theory Ser. A*, vol. 115, no. 1, pp. 105–120, 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.jcta.2007.04.005>
- [16] T. Trung, "A tight bound for frameproof codes viewed in terms of separating hash families," *Des. Codes Cryptogr.*, vol. 72, no. 3, pp. 713–718, 2014. [Online]. Available: <http://dx.doi.org/10.1007/s10623-013-9800-0>